DATA ACCESS TECHNOLOGY FOR THE WARRIOR
COE DATA SHARING TECHNIQUES

Mr. James Pipher, Mr. Stanley O. Davis, Ms. Toni Weir
COE Data Engineering
pipherj@ncr.disa.mil
davis2s@ncr.disa.mil
weirt@ncr.disa.mil
Dr. Glenda Hayes and Dr. Robert Daniels
The MITRE Corporation
ghayes@mitre.org
danielsr@ncr.disa.mil

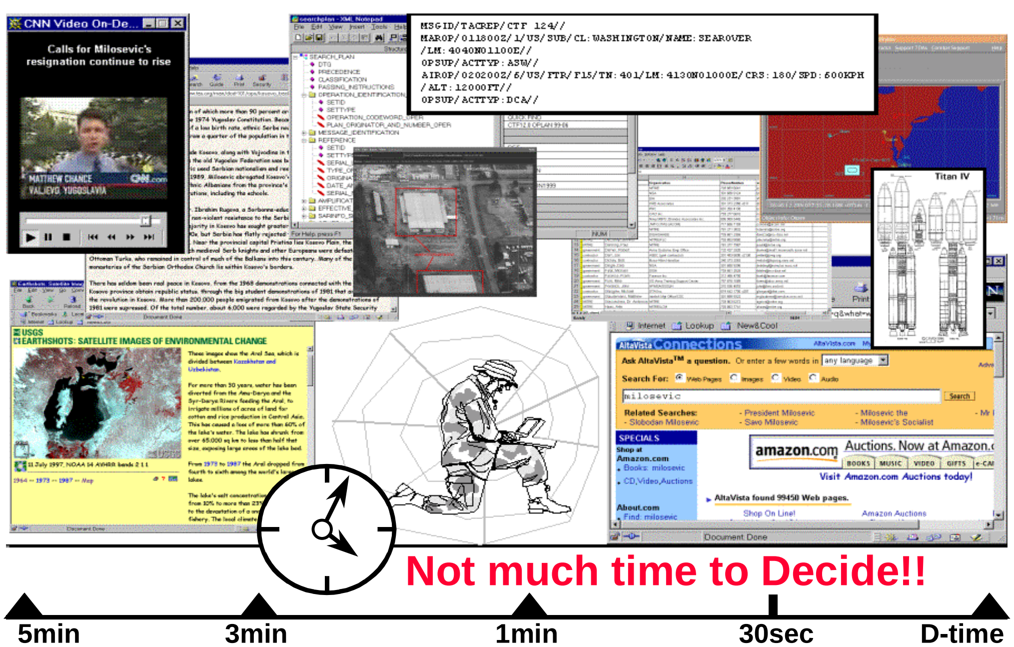
01 December 1999

1. INTRODUCTION

Recent technological advances are helping to produce and consume data at a phenomenal rate and to rapidly transfer data any place on Earth, yet we still lack the ability to chain available data (observations, events, and facts) in an effective manner to optimize decision-making and decision confidence. If we cannot produce, access, and manage data more effectively, we will fail to leverage the DoD's considerable investment in persistent data stores and broadcast streams. If we cannot consistently interpret data, we risk lives and millions of dollars. A recent Mars probe was lost because of a common data failure, confusion reflected in software between metric and English measurement information.

As illustrated below, today's warrior is confronted with an astonishing quantity and variety of data. Most information our warriors need exists *SOMEWHERE*. Not all, but a significant amount is showing up on the Web, the Internet itself and in Web servers on classified networks. New data access technologies are giving every organization "equal opportunity" to post and retrieve massive amounts of information, and this process begets MORE information. Consequently, there is a growing volume of data in lots of different formats (text, spreadsheets, multi-media), and warriors have to separate the wheat from chaff. People often suggest that the answer is simply to provide the warrior a browser. This is not the case.

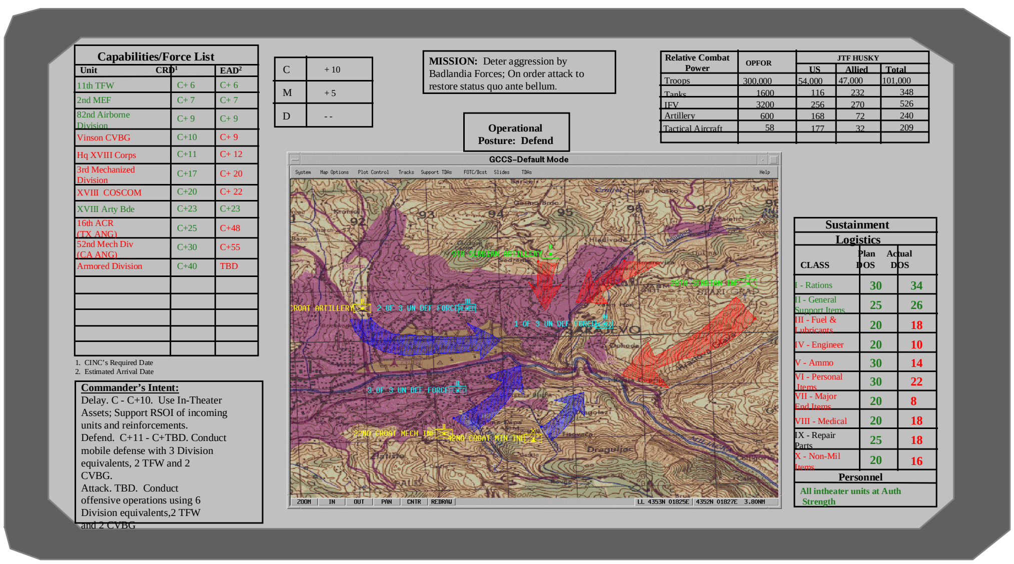
Warrior Dilemma: **Lost in a Web of Data**



Browsing for the right data today requires well structured queries, several search attempts, and plenty of "mouse clicking" that takes time and energy away from warriors, often at a moment when they can least afford it. The AltaVista window depicted lower right above, shows 99,000 hits against "Milosovic." Where does the warrior go from there? This search and refine process can be a costly and fatal distraction. Furthermore, sheer volume may disguise a critical lack of access to correct data. Second and third hand (or worse) data may look like the "right stuff." The warrior needs accurate and complete data FAST -- clear choices from a reliable selection listed quickly and logically, and easily understood.

Smart warrior-computer interfaces are critical to decision support in combat. They are *WHERE* Defense information "rubber meets the road," *BUT* they are by no means the entire vehicle. Below is an example of a modern warrior's screen with an electronic battlemap in the middle and corresponding logistics or status of forces information in tables above and to either side. Displays like this give the Warrior top-level situation appreciation and Indications and Warnings (I&W) capabilities, "at-a-glance." It's a starting point; however, warriors want to "drill down" for much more detail.

Applied Data Access Technology can deliver: Drill-down from Maps, Status Boards or Text



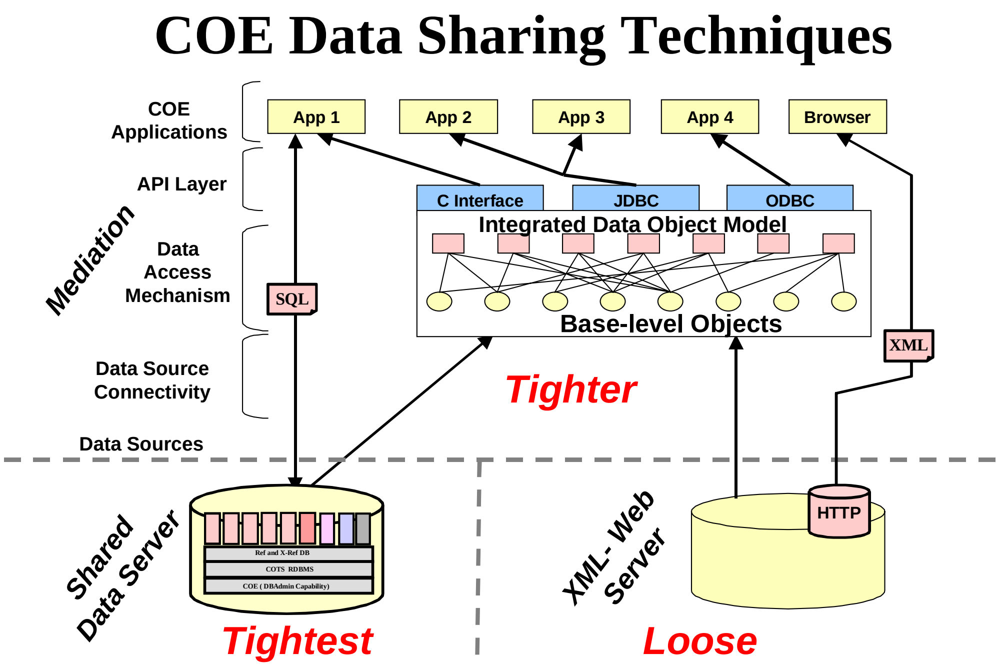
Warriors' Table of Contents
("at-a-glance" decision displays)

Reliable "data food chains" must be created that can efficiently feed warrior processors what they need. Getting top-level data to populate a screen like this is hard enough! Critical data below this level must be painstakingly "rooted out" from many layers of conflicting data structure and semantics in multiple dissimilar data stores or streams physically distributed around the globe. Warrior applications have to know exactly where on the network required data is, how it is labeled, and what form it's in. Any incompatible data must be translated or transformed. Finally, responsive decision support often requires selection and aggregation of data, in essence, a "just-in-time" report. The warrior simply can't do all this data locating, collecting, and processing himself. In the face of increasing volume, the capability to do these tasks must be pre-engineered to make information available on-demand. This is not peculiar to the military; efficient effective data access is also a hot topic in the commercial world where time means money. Like industry, the Defense Information Infrastructure Common Operating Environment (DII COE) is aggressively pursuing new data access technologies in response to warrior needs.

Current and emerging data technologies, such as database gateways, virtual databases, data warehouses, Common Object Request Broker Architecture (CORBA), Java objects, or eXtensible Mark-up Language (XML), enable data sharing from a technical perspective, but do NOT ensure that the exchange of bits between machines has produced valid, reliable information. In addition to technological developments that enable sharing bits, sharing information requires two things: (1) a consistent interpretation of data between sender and receiver (which can be accomplished efficiently only with a common representation that producers and consumers agree on), and (2) a visibility service to make data consumers aware of what data is available from what sources and where in cyberspace those sources are located. Without a common representation and a visibility service, every new project that wishes to exploit these emerging technologies must re-invest in determining the common ground between data consumers and producers in terms of the meaning and location of data.

Within the COE, the SHAred Data Engineering (SHADE) effort provides a comprehensive strategy for mitigating these problems. It includes data sharing approaches, data storage and access architectures, reusable software and data components, development guidelines, and standards for data service developers. SHADE's overall objective is to enable migration of the DII from many redundant, dissimilar, but overlapping, data stores to standardized COE-compliant data services built from "plug-and-play" components that blend multiple data technologies. To do this, the DII COE Data Engineering organization provides engineering support services for system developers and administrators that are intended to reduce the barriers to interoperability. These services include the organization and publication of existing components to encourage reuse. These services also encourage the migration from application-centric data stores to data servers built from common components and extended to meet application-specific requirements.

The DII COE has engineered three approaches to improve warrior data access that exploit new commercial technologies combined with some government-sponsored development efforts and well focused data management. These approaches can be mixed and matched to answer warrior data needs in various circumstances. This paper addresses the data issue by describing these three data sharing techniques, the challenges and opportunities each provides, an example of each technique, and how the COE makes engineering artifacts of each type available for reuse by developers.



2. COE DATA SHARING TECHNIQUES

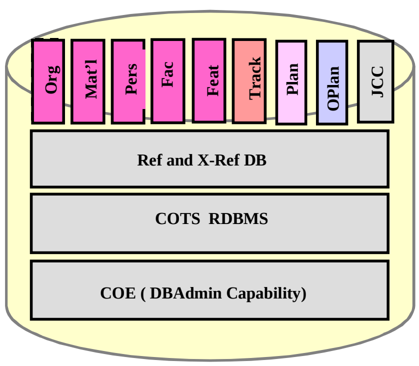
It is impractical to expect all COE systems to adopt a single data sharing approach. Rather, SHADE has identified three compatible data sharing techniques and provides guidance for selecting the one or more to answer a given data access requirement. These techniques are (1) database segments on a shared data server, (2) a mediator or what might be called a "virtual warehouse," or (3) an XML "portal" (each technique will be explained in detail later). Regardless of which technical alternative is selected, a common set of semantics (common representation) must be provided to the calling applications.

SHADE provides the common representation as reusable data engineering artifacts, packaged as segments, that the COE Installer can load automatically and implement immediately. This provides cost avoidance to those programs that reuse the segments, enabling the DoD to spread the investment over multiple programs. In addition to the monetary savings for the program manager (and thus to the taxpayer), reuse of segments that are operational results in greater data interoperability which yields more tailored and better quality information to the warrior. The third benefit from reusing these operational data engineering artifacts is that systems can be developed faster, because COE developers can implement the tested and documented data structures and sample data as is and focus their resources mainly on mission-specific functionality. Therefore, the old adage of cheaper, faster, better – but you must choose which two of the three you want -- is invalid. SHADE segments can make system development cheaper, faster, AND provide better interoperability!

2.1 Shared Data Servers

The shared data server responds to a requirement for tight coordination between users. In this case, the data source is dynamically accessed or the data is replicated to multiple sites (generally as a precaution for a data communication failure or to improve throughput). For replication to work most efficiently, the source and destination data structures must be identical. For multiple users to be able to reuse identical schemas, the semantics (data definitions and business rules for updating and interpreting instance data) must be coordinated among all potential users. If Application A permits Application B to access Application A's data in a read-only fashion, it is important that Application B fully understands the business rules that Application A has implemented, or has assumed. For read-only access, the consequences of improper application of business rules by Application B only affect Application B. However, if Application A permits full data privileges to Application B (create, read, update, delete), Application B can corrupt the shared data source, thus affecting the lives of those depending on its data quality.

Shared Data Server Architecture



- **Standard Server Components**
  - –Kernel
  - –COTS DBMS Segment
  - –References and Cross-references segment
  - –Joint Common Catalog Segment(s)
- **Mission Components**
  - –One or more JC^{2}DS
  - –Mission-specific database segments

Example: Intelligence Shared Data Server (ISDS)

The required coordination can be a significant effort. The ISDS is the relational repository for the Global Command and Control System's (GCCS) Integrated Intelligence and Imagery (I3) capability. The premier database segment loaded on ISDS is a segmented version of Modernized Intelligence Database (MIDB) schema. The MIDB is managed by the Defense Intelligence Agency (DIA) and considered an authoritative source for general military intelligence. DIA sponsors semi-annual meetings involving scores of developers to review proposed changes and identify version releases for accepted changes. Also, since all the involved programs, which reflect an information Community of Interest (COI), are attempting to synchronize on a common representation, any change will inevitably affect many programs and often require expensive adjustments to applications. However, the benefits of this data sharing technique can be equally significant! COE segment packaging ensures that applications or schema will not conflict with one another in the runtime environment and that conflicts will not occur during the installation process. The segmented version of MIDB is loaded (by the COE Installer) on numerous servers in a range of facilities, including CINC command centers, air wing headquarters, and command-configured ships world-wide. Additional applications, such as those found in I3, can leverage data available in MIDB to amplify the Common Operational Picture (COP) used by Joint and Service C4I systems. This provides a wealth of additional (directed and relevant) information for warrior decisions, available with the click of a mouse, yet requiring no transformations that are costly to develop and maintain. In the near term, these shared data servers will capitalize further on their identical schema by enabling Database Replication to distribute and synchronize intelligence information.

2.1.1 Competition

The COE has attempted to constrain the number and variety of Database Management Systems (DBMS) products that are designated as COE-compliant while allowing a range of choices. The current list includes Oracle, Informix, and Sybase, with Microsoft SQLServer coming on board soon. Each of the products conforms to the Structured Query Language (SQL) 95 Standard and has been evaluated against a list of software requirements specified by the COE's Data Access Technical Working Group.

2.1.2 DISA Implementation

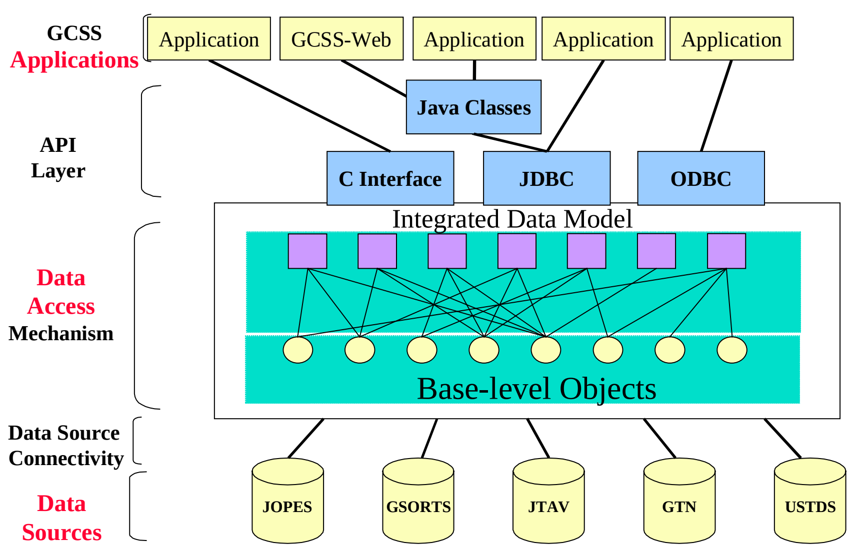
COE provides a clearinghouse for database segments. Each segment is a package of schema (data definition language (DDL) script), models, documentation, and sample data that aid application developers in understanding the structure and business rules for manipulating and interpreting the data. When the coordination on a segment or set of segments expands beyond a single system's boundaries to larger and larger COIs, there should be a corresponding decrease in effort required by others to reuse the data or data structure. The COE's Integration and Runtime Specification (I&RTS) provides metrics for assessing data interoperability of these persistent data schema segments.

2.2 Mediators (Virtual Warehouses)

The technology exists to retrieve and manipulate data from multiple heterogeneous sources. There are components that provide the transport protocol to specific DBMS brands (e.g., Oracle, Sybase). There are drivers that carry the SQL request to the DBMS (e.g., Object Database Connectivity (ODBC), Java Database Connectivity (JDBC), Oracle native driver). There are DBMS-unique network identification configuration files (e.g., tnsnames.ora, interfaces, ODBC administration). There are subtle dialects of SQL that are DBMS and ODBC vendor-specific (e.g., double quotes versus single quotes, symbol for the wildcard in "like" statements). There are protocols for mediating data from one representation to another (e.g., feet to meters, spherical versus ellipsoid, position to latitude/longitude). There are languages to express the algorithms that mediate data (e.g., Java, Smalltalk, VisualBASIC, C++). There are formalisms to represent the integrated model (e.g., object modeling). There is tool support for the modeling formalism (e.g., Universal Modeling Language (UML)). There are techniques and tools to manage the integrated schema. There are techniques and tools by the application developer to use the integrated schema to retrieve data in the runtime environment (e.g., JDBC, C API, ODBC, Java). The capabilities represented by all these services require a team with full breadth and depth in system administration, security, database administration, object modeling, application development, and subject matter expertise (to develop an accurate object model).

Sometimes, it is impractical to reengineer a data store and/or the applications operating on that data store to accommodate additional users and their requirements. Additionally some applications require access to multiple heterogeneous data stores. In these cases, some data transformation or manipulation will be needed to allow new applications to effectively use existing data sources. It may be practical to use mediation technology that can dynamically access a source, retrieve the data, transform the data as required and present the data to the calling application. The mediator acts as a server to calling applications, but acts as a client to data sources. Hence, the architecture is consistent with an N-tier approach.

Mediation Services for GCSS



When the data is cleansed and/or reorganized and then cached (kept in persistent storage) locally, the resulting data store is considered a data warehouse. When the data is retrieved, transformed and/or aggregated on demand, this is considered a virtual data warehouse. The decisions whether to cache data and how much data to cache should be based on criteria such as required query performance, degree and frequency of meditation, and graceful degradation under situations involving loss of communications.

Example: Global Combat Support System (GCSS)

The Global Combat Support System (GCSS) chose to use a commercial mediator technology to implement a virtual data warehouse. This virtual data warehouse is currently known as the Combat Support Data Environment (CSDE). The CSDE acts as the data source for a traditional client/server application - namely the GCSS Combat Support Enhanced (CSE) COP. The CSDE is also the data source for the GCSS portal, a web-centric product designed to supply similar data to thin (web browser) clients.

The CSE COP provides a map-based display of sites (facilities and features) and tracks. Tracks may be associated with units. The association of a track to a unit allows for the use of the COP drill-down query tool. The tool allows a user to navigate through the virtual warehouse, discovering data through directed browsing. For example, a query about an airfield might yield a list of units scheduled to arrive. A subsequent drilldown on one unit might return an Operations Plan (OPLAN) link, which would in turn lead to other units and or locations associated with an operation. The key is that the virtual warehouse has navigation paths defined that will return useful information for a user who doesn't know what to ask, but can recognize what he needs when he sees it.

2.2.1 Challenges

Mediator data access requirements are typically in the form of a question or query that a potential user needs to have answered. Often the user with the requirement is unable to describe the best (or any) source for the data. Therefore, data analysis has to be performed to investigate potential sources of new (or better) data. Once a source is identified, a memorandum of agreement (MOA) must be negotiated to acquire read-only access to each of the data sources. This involves administrative issues such as accounts, access control, security and audit trails. It may also involve arranging to move data from an unclassified source to a classified environment. The coordination with the sources must be an active process at the operational level. Any scheduled outages, schema changes, or infrastructure changes have the potential to cause a virtual database data access to fail. In addition to structural metadata, source system business rules may be critical to the proper retrieval, interpretation, and presentation of the data to a user. Since many legacy systems have these business rules implemented via applications rather than via DBMS mechanisms, there may be no automated method for capturing this information.

As challenging as this process for direct access to a data source seems, many data sources are themselves data warehouses in varying degrees. Two examples are the Global Transportation Network (GTN) and Joint Total Asset Visibility (JTAV). GTN and JTAV processes and aggregates dozens of other source systems. Even a relatively direct source such as the Joint Operations and Planning Execution Systems (JOPES) has extensive data updates from service feeder systems and has a large number of tables, which contain airfield and seaport data that should come from an authoritative source (e.g., National Imagery and Mapping Agency (NIMA)).

Given a detailed understanding of available source data, data integration faces the challenge of determining which data elements in which systems allow for navigation between systems and the consolidation/integration/aggregation of query results from multiple systems. Once virtual views have been defined that combine data from multiple sources, the data integration team needs to document the mediation and manipulation that is performed by the mediator. If this knowledge is only available from examination of the mediator application source code, then we have simply created a more sophisticated legacy application. It may not be possible for the developers of any data warehouse (virtual or not) to completely understand the business rules that apply to all of the source systems data. It should be possible, however, to document an interpretation of source system data via the warehouse in such a manner that a person who does understand the business rules for that source system will be able to recognize incorrect interpretations and inappropriate manipulations.

2.2.2 DISA Implementation

Mediators are often a set of COTS products that have been integrated and supplemented by another vendor. The data access mechanism, provided by a major RDBMS vendor, includes utilities for capturing source system metadata and managing source system account and password data. Some source systems, such as the JOPES and the Global Status of Forces, Readiness and Training System (GSORTS), may provide full access to their schema, while others such as GTN provide only limited interfaces in the form of stored procedure queries.

The mediation mechanism is a meta-catalog that contains base-level objects, which correspond to source system data objects such as relational tables and views and contains virtual views that provide combinations of data from multiple sources and/or derived or reformatted data from a single source.

The DISA approach uses Rational Rose to create a UML model that represents the virtual views of the meta-catalog as classes and the virtual attributes as class attributes. The virtual attributes are mapped to attributes in the base-level objects (data source columns). The UML model is read by a parser program, which generates the code required to update the meta-catalog to match the model. Therefore, the UML model provides a non-proprietary description of the virtual data warehouse's content. It is anticipated that the UML model will also be used to generate Java components and XML components in future versions.

2.3 XML Web Portal

The third technique for data access provided by the COE addresses requirements for loose coordination between data producers and data consumers. "Web portals" provide a window into databases by exploiting the internet to publish selected data via web servers. Data is published via documents generated on demand by databases and read by internet browsers (Netscape and Internet Explorer (IE) 5.0)) or suitably enabled applications. Document contents are annotated with the new internet standard XML. Annotated data in XML documents is easily extracted and made available to end users or applications. Since standard commercially available web servers are needed to publish XML documents, less design work is required to gain access to the database. Hardware and software components are relatively cheap and available off the shelf.

2.3.1 XML

For centuries, editors have marked up manuscripts with annotations to indicate to printers the appearance and style of documents they will produce. Much of the power of the World Wide Web comes from annotating documents with a fixed set of tags which tell a browser how to display a document or how to hyperlink one document with another. This is done through a simple HyperText Markup Language (HTML). XML follows the principles established for HTML with the added feature that document originators can create their own tags. This added feature permits the originator to identify the meaning to contents of the document. This example, <zipcode> 22136 </zipcode> , indicates that the string of characters “22136” represents a zipcode, not an annual salary. In addition, XML documents can contain a description of their composition in the form of a document type description (DTD). Tags and DTDs, in effect, make a document and its contents self-describing, thus rendering their interpretation mechanical (and deterministic). The bottom line is that data can be published in a standard way where producer and consumer need only agree on the meaning of tags in order to share data.

2.3.2 DISA Implementation

With the benefits that XML can bring, there are also significant challenges that COE Data Engineering anticipates. There are extremely low entry barriers to using XML. The tags and XML documents can be created with mere text editors. The result is likely to be proliferation of non-interoperable tags. Desired, recommended tags need to be collected, validated and made visible and accessible for COE developers to reuse. What is needed is a simple, flexible, affordable, domain-focused approach for tag and DTD management. The feedback from XML developers affiliated with the COE is that they want to use common tags, however they need an easy mechanism for finding "good" tags quickly.

To address this, the COE has implemented an XML Registry to record agreements on the meaning of XML tags. The XML Registry will provide tag guidance for COE developers in a web-enabled manner. In order to disambiguate tags, they need only be unique within the context of a namespace. The COE is designating a set of approved namespaces. A manager will be appointed for each namespace who will control the assignment of tags within the namespace. By COE CRCB direction, systems that use XML as a public interface must register XML tags in the Registry within an appropriate namespace. In order to interpret a document, the appropriate tag need only be referenced in the COE XML Registry. The Registry facilitates control of tags and encourages their reuse. SHADE has initialized the Registry with tags derived from the Joint Command and Control Database Segments (JC2DS) which are being packaged as change proposals to the Data Standards. This synchronization with JC2DS ensures interoperability between the relational database format and the XML format.

Example: Track and Relational Data Synchronization

One very significant place the COE is using XML today is in our data synchronization project to establish links and resolve the discrepancies between data from incoming messages on near real time activity in the battlespace and less volatile more detailed data contained in "source" databases. At the heart of this design is our cross-reference capability that we hope will provide warriors down at the foxhole level with access to important amplifying data as well as solve some redundant COP display problems. This synchronization project features an "Archiver" kit that allows developers with Relational Databases to subscribe to Track Management System (TMS) track update events via XML and load them into persistent data stores where they need to retain track histories. The more robust TMS update function allows the holders of tactically significant data in relational data stores to actually start or update existing tracks while retaining a link back to detailed data in RDBMS records.

2.3.3 Commercial Products

Commercial products that exploit XML are rapidly appearing on the market. The latest releases of the major relational DBMS include (or will include) the capability to generate XML documents in response to an SQL query. The Microsoft Office 2000 suite uses XML as a native file format. Microsoft IE V5 can be easily programmed to perform sophisticated manipulations on XML documents.

In conclusion, XML Web Portals form a flexible, relatively inexpensive technology to introduce data sharing capabilities to existing databases. An extensive white paper on XML is under development by SHADE for release during 2000.

3. COMMON REPRESENTATION

Key SHADE objectives are to (1) leverage investments in existing databases, data structures, and data values, (2) promote interoperability through their reuse, and (3) provide a foundation for data fusion. A prerequisite to achieving these objectives is a common representation of battlefield data. The common representation provides "to be" migration objectives, a common understanding of the data, agreement on core objects, their identifiers, and valid domain values. Additionally, it constitutes the core set of battlefield data which mission applications extend as required. The common representation is maintained as a logical model, but is manifested in multiple physical forms (e.g., Informix, Sybase or Oracle databases, XML documents, Flat files, Object Oriented DBMS (OODBMS), etc.). The common representation is being evolved by the COE Chief Engineer's Data Engineering team from the existing C2 Core Data Model, a subset of the Defense Data Model (DDM), and from data structures/semantics used by key C3I systems. It is being made available as COE component database segments, XML tags/metadata, reference set code values, and other forms as required. These and other COE data products can be located via the COE's Data Emporium (http://diides.ncr.disa.mil/shade).

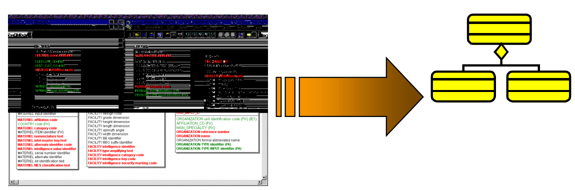
The COE Data Engineering strategy is to foster inter-community technical coordination to resolve data representation and sharing obstacles to incrementally develop a common representation for the core battlefield objects. A common representation is an agreement among data providers and data consumers about the labels, definitions, and data types of objects that are to be shared, as well as the relationships among those objects. The common representation is valid within a defined community of interest. A common representation could take the form of a logical data model or object model. If organizations do not share some common entities and attributes, the combining of data from different sources loses meaning since alternatives cannot be compared on a similar basis. Furthermore, to combine data from multiple sources, there must be a basis for guaranteeing the unique identification of all data elements, otherwise data quality is seriously compromised. If organizations do not share the same primary key attributes (unique identifier) for a shared entity, it is problematic to ensure that double counting hasn't occurred. Rather than tackle all possible entities and attributes that might occur within the DoD, SHADE is fostering an incremental improvement approach by initially tackling specific major battlespace objects that occur across all services and the intelligence community, most of which will directly affect the ability of the warfighter to perform his objectives.

SHADE is brokering technical exchange meetings for programs that have a need to share data but have been unsuccessful in arranging agreements in the past or are unaware of significant data interdependencies. These meetings, which involve the engineers responsible for major Command and Control and Combat Support systems data service development, result in sharable, plug-and-play products. These data products are the basis for interoperability within a given version of the COE. Their structure and semantics are the foundation for engineering change proposals, where required, to the DOD Data Standards.

Common Representation

Consistent Interpretation of Data

- **Common logical specification, same semantics, common value set, incremental Standards refinement for primary Battlefield Objects.**
- **Used for:**
  - –Persistent data storage (databases) *and*
  - –Data Integration (via neutral schema)
- **Structure can morph to multiple data and metadata formats**
  - –RDBMS
  - –Flat File
  - –UML
  - –XML
  - –Interface Definitions

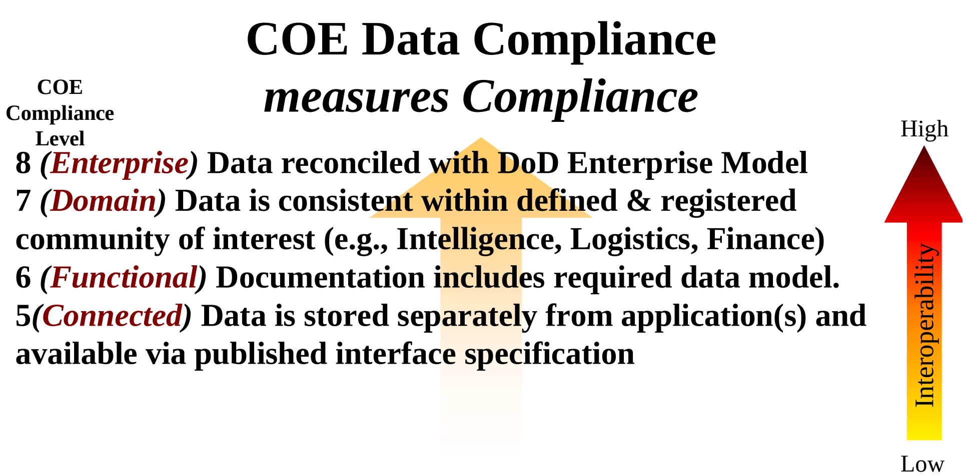


With the advent of newer data technologies, it is crucial to recognize that a common representation will have many physical manifestations, but that the semantics must be able to bridge peculiarities of each form. Among the many possible forms the logical model might take, we immediately recognize the IDEF1X data model, relational database form, an XML DTD, and UML object model.

Recently, the re-engineering effort for JOPES indicated a commitment to support the development and reuse of the JC2DB Org segment. The developers for the COP, which produces a correlated, fused, and geospatially-referenced display of objects of interest to multiple decision makers, have also acknowledged the importance of the common representation to facilitate the sharing of data.

3.1 Data Interoperability Metrics

Since SHADE doesn't “own” data, it must provide guidance to data owners for improving data sharing. SHADE has recently refined the I&RTS (the Rule Book for COE Compliance) to assess the sharability of a system's data. Joint Pub 1-02 states that metrics need to be identified to properly assess the level of interoperability. SHADE defines sharability as the level of effort needed to reuse data properly. The more exclusive a data sharing group is, the more effort is needed to reuse the data.



In Levels 1-4, data has not been separated from the application. Levels 5 and above, therefore, deal with persistent data stores. In Level 5, the ability to share data requires significant effort and is probably point-to-point agreements. Since there is a public interface to the data source, however, the data is sharable. The current segmentation process is still appropriate for Level 5.

In Level 6, the database segment also contains a model for the data (either data or object model) that exhibits the same characteristics as a fully attributed, normalized logical data model. The segment must also contain a machine navigable cross-reference between the model and the physical schema. To enable systems to share data, producers and consumers must agree on the name, description, data type, data size, and constraints. The normalized model reflects the business rules (or underlying assumptions) that are either being implemented by the DBMS (if there is a one-to-one correspondence between the elements in the logical model and elements in the physical schema), or must be implemented within the applications writing to the database (if the physical schema has been de-normalized). If a system is going to permit other applications to read its data and the applications do not properly interpret the data in the system, it may negatively impact the data quality of the new application. However, if the new application is given create, read, update, and delete, the new application may seriously compromise the data quality of the system's data. Therefore, it is imperative that when a system grants access to its data, there is every attempt to provide a current, active model. For Level 6 compliance, there has been no coordination outside the immediate system on the logical model.

In Level 7, the requirements build on Level 6 and indicate there has been coordination within some identified and registered COI. A COI is loosely defined as a group with a current, operational data sharing imperative. The coordination is reflected in the logical model that accompanies the segment.

In Level 8, the COI is the DOD Enterprise, thus the coordination is optimal. In addition to these refinements, we factor in deviations from the norm (whether it be Level 7 or Level 8) by stating that a segment may be compliant if it provides the mediation necessary to share data via the logical model specification. This shifts the burden of developing and maintaining mediation to those who deviate from the norm. Then, it becomes much easier to assess how expensive the customization really is.

Finally, the COE's engineering approach to data prescribes techniques for identifying, deconflicting, and resolving conflicts among data standards, database segments (which contain operational data schemas), and message standards. XML is a text-based, data interchange format, the United States Message Transfer Format (USMTF) message community has decided to migrate to that format. Commercial-off-the-shelf (COTS) DBMS vendors and Java component providers are providing the functionality for DBMSs to ingest and output XML. Thus, we have a situation where the same data format (XML) can be used by all three. Now we may have a conducive environment for more easily seeing the overlaps and resolving redundancies. The COE feature will also allow us an opportunity to synchronize database segments, XML tags, and community models with the COE versioning cycle.

3.2 Benefits of a Common Representation

The benefits for the Warrior are

- Improved Interoperability through COE Compliance
- Better C2I /Combat Support data availability, accuracy and consistency
- Intelligent search and retrieval

- Efficient warrior access to authoritative tactical data resources
- Greater warrior access to non-tactical data inside *and* outside COE
- Rapid warrior access to fused dissimilar data types
    - imagery
    - general purpose intelligence
    - near real time track data

The benefits for the Taxpayer are

- Reduced expenditures for basic data store development and maintenance
- Lower interface development and maintenance costs
- Fewer data management and data access hardware-software suites

4. JOINT COMMON CATALOG

Next, the common representation must be capable of being engineered and generated into multiple physical forms; including database schemas, UML models, and XML DTDs (or DCDs). The common meta-object facilities ((MOFs) - OMG's term for the services to construct, populate and maintain a metadata repository) provide the services necessary to engineer the common representation and transform the common representation into the multiple forms.

Making the operational data visible, queryable and retrievable via the terms expressed in the common representation is essential to achieving data interoperability within the COE being fielded by the DoD. These visibility, query, and retrieve services (which we associate with a "runtime MOF") raise the importance and utility of the common representation out of the category of paperweights and into the realm of facilitating information retrieval and will be included within the COE Data Engineering infrastructure.

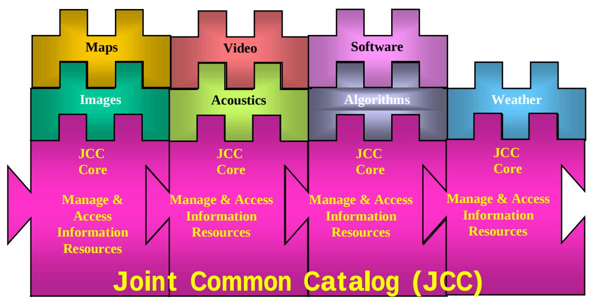
4.1 Identifying authoritative data sources

There is no list that identifies the authoritative sources for various types of data. The CSE-COP project had incorporated airfield and seaport data from the JOPES database because that source was available. Later, the team found the same concepts represented within a NIMA source and discovered the data within JOPES was not being maintained and, consequently, was seriously outdated. When a data source is not widely known or it becomes complicated to negotiate an access agreement, less authoritative sources will be substituted and the resulting decision quality can compromise the warrior's safety or the mission's success.

4.2 Common Visibility - Joint Common Catalog (JCC)

Joint Common Catalog:
Run-time Librarian Services

What is it? Where is it? How does the Warrior get it?



Provides Run-time Visibility of Information Resources
to enable Data Access as required by COE-compliant Applications

To leverage the DoD Enterprise memory, the COE needs a common visibility service for data. Data (or any information resource) cannot be reused if its existence is not easily recognized. Some time ago, Defense Modeling and Simulation Organization (DMSO) asked SHADE to host a meeting among various groups (METOC, NIMA, DMSO) that are actively engaged in developing web-based catalogs to manage discovery and access to their information products: Master Environmental Library (MEL) to weather and oceanographic data, United States Imagery and Geospatial Information Systems (USIGS) to imagery, and Modeling and Simulation Resource Repository (MSRR) to models and simulations. The inefficiencies and non-interoperability of stove-piped catalog development were cited as the driving concern among the organizations. SHADE was contacted because the COE provides the tools and configuration management for packaging and reuse that allows organizations to leverage one investment over many implementations (cost avoidance) AND achieve interoperability by using a distributed approach to catalog interaction.

The JCC is a SHADE initiative to define the services and schema requirements that are common across all types of information resource catalogs and package the services as application and database segments so that they can be reused in the COE fashion. These segments would be available as components from the COE "shelf" to be installed in COE systems and could be supplemented by "plug-in" components that deliver specialized requirements unique to certain types of information resources. The "core" component would also have the capability to participate in a federation of information resource catalogs by complying with a pre-defined set of federation rules.

An installed JCC can be used at build time to discover information resources that are available in the run time environment. It could also be used at run time as a buffer between staggered application and database development cycles. In COE 4.x, a JCC will be standard on COE Data Servers to provide ready access to logical model information and mappings to the physical schema implemented in the local data store (currently a relational database). This coordination between build time (modeling activity) and run time (applications accessing data) provides the rules and conditions to other developers for properly reusing a data source. Since the model becomes the basis for data discovery and access, the currency of the model can be validated with queries that retrieve operational data. Only when the model becomes an active component in data discovery and access will there be higher assurance that the model accurately reflects the business rules to effectively use the data source.

4.2 Self-describing Data Servers for the COE

COE Data Engineering will provide an information resource catalog, JCC, as a standard component for every COE data server. The database install will extract the mapping of the logical model to physical elements from a COE Level 6 or above segment and insert these mappings into the catalog. The catalog would provide navigation and search capabilities of holdings on the data server via their logical model, as well as actual data retrieval. The catalog could be used by applications as a means of insulating themselves from changes to the physical data structure. Applications would still have the option of replicating the access information locally and directly accessing the data via the physical structure, but risk having their applications break when changes occur to the schema.

Therefore, the catalog would provide a uniform, lightweight approach to querying the contents of each COE server, from both a logical and a physical perspective. The catalog resident on each data server would be able to interact with catalogs on other data servers in a loose federation. The model information would be exchanged between requesting application and catalog using XML and data would be returned to the requester as XML, both viewable with the next generation browsers (two products already available in beta). The catalog would be capable of introspection, thus revealing its capabilities (methods) to a requesting application. The catalog would be accessible via Java and CORBA, at a minimum. The catalog would also provide a subscription capability to control access to logical models and physical data by associating subscribers with roles. Applications could direct their data request to the catalog, which would then access the source and return the data to the application. Since two (Oracle and Sybase) of the three approved COE DBMS products have announced XML as a prominent part of their output strategy, the catalog should be able to use this advantage to standardize the data return format to the requesting application.

The catalog can be the mechanism for standardizing data access security for each of the data sources, using pre-established, shared data source username and roles within the data source's system and identifying the actual username and correspondence to pre-established username.

5. CONCLUSION

Through modern data access methods COE Data Engineering has started to provide reliable "data food chains" that can efficiently feed warrior processors what they need. A first version toolset for rooting out data "gold" from many layers of conflicting data structure and semantics in multiple dissimilar data stores is nearing implementation. Warrior applications built on the COE in 2000 and beyond will know where on the network the data they require is, how it is labeled, and what form it's in. They will have the required data translated or transformed for them, and where necessary selection and aggregation services will formulate "just-in-time" reports. The payoff for the warrior will be a significant reduction in the burden of locating, collecting, and processing data manually or with little automated assistance.